Linkage Disequilibria and the Site Frequency Spectra in the *su(s)* and *su(w^a)* Regions of the *Drosophila melanogaster X* Chromosome

Charles H. Langley,* Brian P. Lazzaro,*,† Wendy Phillips,*
Erja Heikkinen*,‡ and John M. Braverman*,§

**Center for Population Biology and the Section of Evolution and Ecology, University of California, Davis, California 95616, †Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, ‡Center for Scientific Computing, 02101 Espoo, Finland and §Department of Biology, Loyola University, Chicago, Illinois 60626*

Manuscript received June 25, 1999
Accepted for publication August 23, 2000

ABSTRACT

Over the last decade, surveys of DNA sequence variation in natural populations of several Drosophila species and other taxa have established that polymorphism is reduced in genomic regions characterized by low rates of crossing over per physical length. Parallel studies have also established that divergence between species is not reduced in these same genomic regions, thus eliminating explanations that rely on a correlation between the rates of mutation and crossing over. Several theoretical models (directional hitchhiking, background selection, and random environment) have been proposed as population genetic explanations. In this study samples from an African population ($n = 50$) and a European population ($n = 51$) were surveyed at the *su(s)* (1955 bp) and *su(w^a)* (3213 bp) loci for DNA sequence polymorphism, utilizing a stratified SSCP/DNA sequencing protocol. These loci are located near the telomere of the *X* chromosome, in a region of reduced crossing over per physical length, and exhibit a significant reduction in DNA sequence polymorphism. Unlike most previously surveyed, these loci reveal substantial skews toward rare site frequencies, consistent with the predictions of directional hitchhiking and random environment models and inconsistent with the general predictions of the background selection model (or neutral theory). No evidence for excess geographic differentiation at these loci is observed. Although linkage disequilibrium is observed between closely linked sites within these loci, many recombination events in the genealogy of the sampled alleles can be inferred and the genomic scale of linkage disequilibrium, measured in base pairs between sites, is the same as that observed for loci in regions of normal crossing over. We conclude that gene conversion must be high in these regions of low crossing over.

THE causes of the empirically observed reductions in DNA sequence polymorphism in chromosomal regions experiencing a low rate of crossing over per physical length (Aguadé *et al.* 1989; Stephan and Langley 1989; Aguadé and Langley 1994; Aquadro *et al.* 1994; Stephan 1994) have been a subject of intense theoretical analysis (Kaplan *et al.* 1989; Charlesworth *et al.* 1993; Charlesworth 1994, 1996; Stephan 1994; Braverman *et al.* 1995; Gillespie 1997). The lack of any parallel correlation between crossing over per physical length and divergence between species firmly establishes that the mechanism(s) creating the association of crossing over per physical length with polymorphism must be operating at the population genetic level (Berry *et al.* 1991; Begun and Aquadro 1992; Martín-Campos *et al.* 1992; Langley *et al.* 1993). Accumulating evidence in *Drosophila ananassae* (Stephan and Langley 1989; Stephan *et al.* 1998), Mus (Nachman 1997), Aegilops (wild relatives of wheat; Dvorák *et al.* 1998), and Lycopersicon (wild relatives of tomato; Stephan and Langley 1998) indicates that this relationship can now be expected in many organisms.

Two dichotomous models based on gene frequency perturbations associated with selection at linked loci were proposed to explain the observations. The first is the extension of Maynard Smith and Haigh's (1974) original "directional hitchhiking effect" analysis to the comparison of genomic regions with differing levels of crossing over per physical length (Kaplan *et al.* 1989). This model supposes that rare (perhaps newly arising) highly favored variants occasionally spread rapidly to fixation in a population dragging with it the genetic haplotype upon which it originated. The size of the region (measured in morgans) affected by such hitchhiking events is of the same order as the selection coefficient associated with the favored, rare variant. Thus in regions of low crossing over per physical length the impact of each selected substitution is much greater. The few polymorphisms at tightly linked sites that have survived or emerged since the last hitchhiking event are likely to be rare, so a skew in the frequency spectrum

Corresponding author: Charles H. Langley, 3342B Storer Hall, Center for Population Biology & Section of Evolution & Ecology, University of California, 1 Shields Ave., Davis, CA 95616-8554.
E-mail: chlangley@ucdavis.edu

is expected under the directional hitchhiking effect (Aguadé *et al.* 1989; Braverman *et al.* 1995). While the original survey of the *yellow-Achaete Scute Complex* (*ASC*) revealed the expected skew (Aguadé *et al.* 1989; Martín-Campos *et al.* 1992), subsequent surveys have yielded equivocal results (Aguadé *et al.* 1994; Braverman *et al.* 1995).

The *background selection model*, proposed as an alternative to the *directional hitchhiking model*, posits that the deleterious mutation rate at closely linked selected sites is sufficiently great that most chromosomes bear one or more deleterious mutations tightly linked to the selectively neutral sites being studied (Charlesworth *et al.* 1993; Charlesworth 1994). Since selection is assumed to be sufficiently strong to effectively prevent these mutation-bearing chromosomes from leaving any descendent copies over the long term, the effective population size of deleterious-mutant-free chromosomes can be reduced to a small fraction of that expected in a selectively neutral region recombinationally distant from a selected locus. Obviously in genomic regions of very low crossing over per physical length the total deleterious mutation rate (per unit crossing over) will be higher. The parameter domain in which the predictions of the *background selection model* predict the observed genomic distribution of the levels of polymorphism is circumscribed by the known variation among genomic regions in crossing over per physical length, the experimentally determined limits on the deleterious mutation rate, and the model's apparent sensitivity to even small levels of crossing over (Hudson and Kaplan 1995; Charlesworth 1996). Hudson and Kaplan (1995) have pointed out that the density of rare (presumably deleterious) transposable elements in regions of low crossing over per physical length in *D. melanogaster* may be sufficient to cause the background selection effect.

In a superficial and imprecise sense both the hitchhiking effect and background selection models are associated with a *reduction in effective population size.* Under the *background selection model* the spectrum of selectively neutral site frequencies for a population of size N is close to that expected for a population of size Nf_0, where f_0 is the proportion chromosomes free of deleterious mutations (Charlesworth 1994). Negative skew in the frequency spectrum due to background selection is "unlikely to be frequently associated with significant values" of the available test statics (Charlesworth 1996). Under the *directional hitchhiking effect*, linked polymorphisms are being episodically "swept out" and then replenished slowly by a combination of mutation and genetic drift. The expected skew in the site frequency spectrum under the directional hitchhiking effect reflects this continuing state of "recovery" of the neutral polymorphism (Aguadé *et al.* 1989; Braverman *et al.* 1995). In his extensive simulations Gillespie (1997) has observed such a skew in the frequency spectrum of neutral sites linked to a selected locus undergoing any number of random-environment-selection processes; all have the tendency to strongly reduce standing variation. Here we report substantial reductions in DNA sequence polymorphism at two loci, *su(s)* and *su(w^a)*, located in a region of low crossing over per physical length near the telomere of the X chromosome of *D. melanogaster*, in large samples of alleles from an African and a European population. Consistent with the directional hitchhiking effect, a distinct skew toward rare sites is evident.

Substantial intragenic recombination is also evident in the inferred history of the sampled alleles. Molecular population genetic models typically fail to incorporate gene conversion. The effects of this recombination process are often assumed to be the same as crossing over or it is assumed that the scales of gene conversion tracts and/or of the rates of gene conversion are negligibly small. Analyses of both the *hitchhiking effect* and the *background selection* models have ignored gene conversion. The high level of historical recombination evident among our sampled alleles appears inconsistent with our expectations on the basis of surveys of loci in regions with normal levels of crossing over per physical length and the known reduction in crossing over per physical length at the tip of the X chromosome. We propose that gene conversion is the likely mechanism of this recombination and discuss its implications for the models proposed to explain the correlation of DNA sequence polymorphism with rates of crossing over per physical length.

MATERIALS AND METHODS

Drosophila stocks and DNA preparation: Genomic DNA was isolated from 50 independent isogenic X chromosome lines of *D. melanogaster* extracted from the population in the Sengwa Wildlife Reserve, Zimbabwe, Africa (Begun and Aquadro 1993) and 51 isogenic X chromosome lines from Barcelona, Catalonia, Spain (Martín-Campos *et al.* 1992). Genomic DNA was prepared as described previously (Aguadé *et al.* 1994).

Single strand conformation polymorphism survey: In an earlier report (Aguadé *et al.* 1994) it was demonstrated that DNA sequence polymorphism could be efficiently surveyed by a stratified approach of first applying single strand conformation polymorphism (SSCP) to PCR fragments, followed by direct DNA sequencing of representative members of each "allelic" SSCP class. The SSCP survey allows for relatively large sample sizes. If the amount of polymorphism per fragment is sufficiently low, only a small proportion of fragments must be sequenced. In genomic regions of low crossing over per physical length in *D. melanogaster*, where expected heterozygosity per site (or pairwise differences or gene diversity) is low (<0.001), surveys of >3 kb for sample sizes of 50 are practical. This method is not practical in genomic regions of normal crossing over per physical length because the number of polymorphisms within each SSCP fragment (and thus the number of SSCP "classes") is typically so large that the amount of direct DNA sequencing approaches that associated with sequencing each sampled allele. Figure 1, a and b, shows the PCR-amplified segments of noncoding sequence, which totaled 3213 and

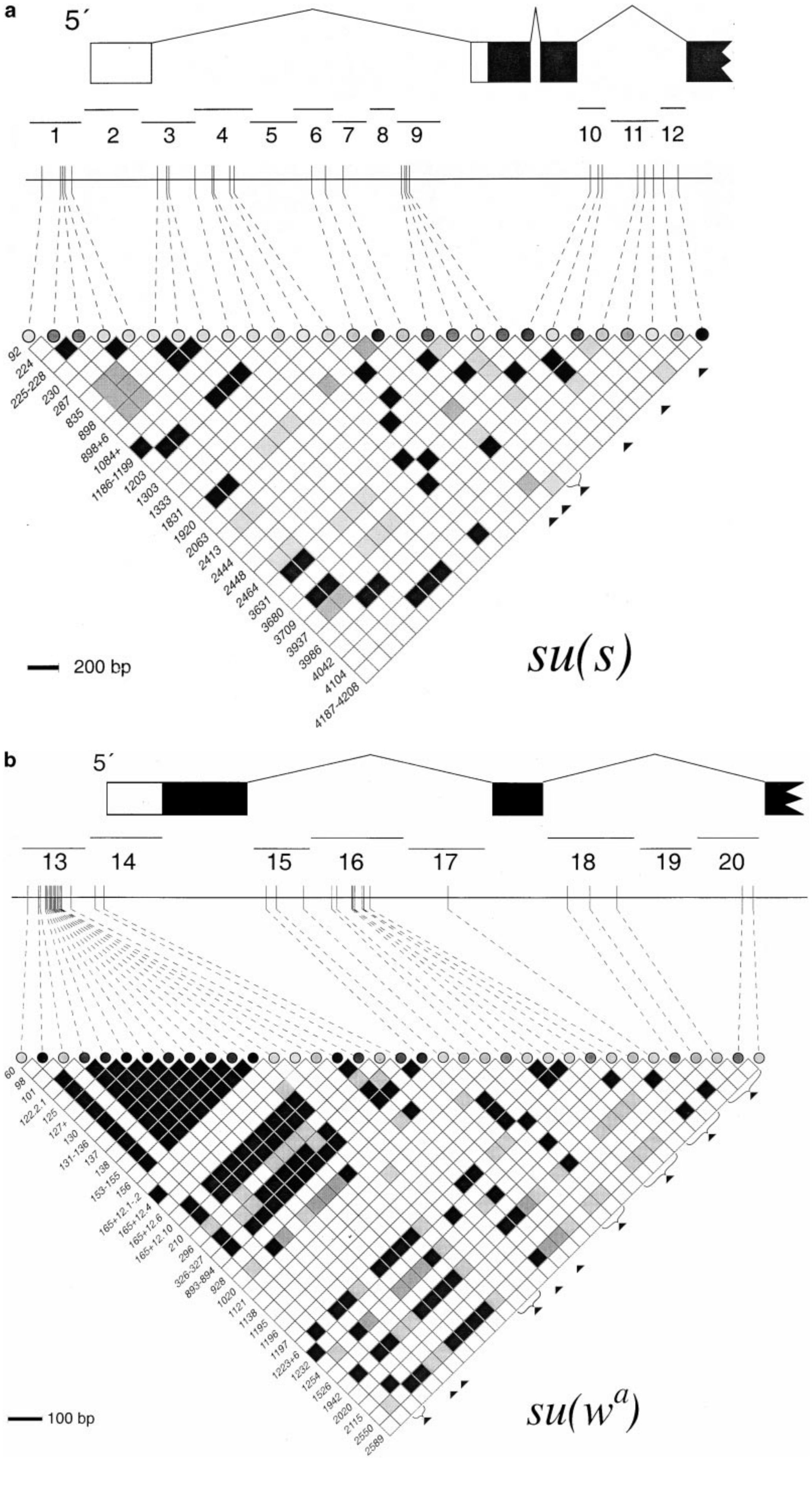

FIGURE 1.—Polymorphism and linkage disequilibria in the African sample. The positions of polymorphic sites detected (ticks on the third level) in the survey of the 5′ portion of the *su(s)* (a) and of *su(w^a)* (b) in 50 alleles from an African population are indicated below a depiction of the gene structure (the open box is the 5′ untranslated portion, the solid boxes are coding portions of exons, and the thin lines are the introns). Only those polymorphisms at which the least common state occurred twice or more in the sample are shown. Directly below the gene structure are the positions and sizes of the PCR fragments surveyed by SSCP and DNA sequencing. The triangular matrix of squares represents the statistical significance determined by Fisher's exact test (uncorrected for multiple tests); white, $P \geq 0.05$; light gray, $P < 0.05$; dark gray, $P < 0.01$; and black, $P < 0.005$. The dashed lines connect the polymorphic sites to their corresponding columns in the matrix. The shading of the circles at the top of each column increases with the expected heterozygosity of the polymorphic site. Along the left margin of the matrix are the positions in the published sequence of the corresponding polymorphic sites. Along the right margin are the inferred positions of the "minimum" number of recombination events in the history of the sample (HUDSON and KAPLAN 1985).

1955 bp for the *su(s)* and *su(w^a)* regions, respectively, not counting the length of the primers and small segments of overlap (Aguadé *et al.* 1994). SSCP fragments surveyed ranged in length from 161 to 358 bp. PCR amplifications, sample loading and electrophoresis, silver staining, and scoring were conducted as previously described (Aguadé *et al.* 1994).

DNA sequencing: DNA sequences were determined as described in Aguadé *et al.* (1994) or using standard protocols on the ABI 377 automated sequencer.

Estimation of DNA sequence variation: Following the estimation procedure in Aguadé *et al.* (1994), the reported estimates of the average number of pairwise differences (per site) were corrected for failure of our SSCP assay to detect all variants. The estimates of insertion/deletion variation were not corrected. Some alleles were not scored for SSCP but were directly sequenced. These SSCP-unscored alleles were treated as separate classes for purposes of estimation of average number of pairwise differences, $\hat{\pi}^*$ (Aguadé *et al.* 1994). $\hat{\pi}^*$ is corrected by $n/(n-1)$ for sampling bias, and bootstrap confidence intervals for sampling, corrected by $n^2/(n-1)^2$, are placed on the estimates of the overall average pairwise differences.

The calculated values for Tajima's D (Tajima 1989) and for estimates of linkage disequilibrium were based on DNA sequences inferred by assuming each sampled allele within an SSCP class was identical in sequence to that of sequenced members. The few within-class variant sites (detected by direct sequencing) were ignored in the calculations of Tajima's D and of linkage disequilibria. If the frequency spectrum of within-class variants is similar to that of detected polymorphisms, then no bias in the analysis of the site frequency spectrum should arise. The confidence intervals for Tajima's D were determined by simulation, as in Braverman *et al.* (1995). The distributions of Tajima's D under various rates of hitchhiking were determined as described in Braverman *et al.* (1995). Estimates of $3Nc$ (where N is the population size and c is the intragenic rate of recombination) were obtained according to Hudson (1987). The "minimum number of recombination events" was determined by the algorithm in Hudson and Kaplan (1985).

RESULTS

Tables 1–4 present the SSCP scoring and the inferred and observed (boldface) DNA sequence state at each of the polymorphic sites in each surveyed X chromosome line at the *su(s)* and *su(w^a)* loci (see Figures 1 and 2). Including polymorphic insertions, a total of 3220 (3213 + 7) nucleotide positions were surveyed at the *su(s)* locus; of these, 112 and 54 vary (as substitutions or as part of insertion/deletion variation) in the samples from Africa and Europe, respectively. Of the total of 1983 (1955 + 28) nucleotide positions surveyed at the *su(w^a)* locus, 83 are polymorphic in the African sample, while 89 in the European sample are segregating either as single substitutions or as part of insertion/deletion polymorphisms.

Table 5 presents the estimated average number of pairwise differences, $\hat{\pi}^*$ (and bootstrap confidence intervals), for each locus and population as well as those from Aguadé *et al.* (1994) for comparison. The African sample harbors substantially more DNA sequence polymorphisms than do those from North America and Europe. $\hat{\pi}^*$ estimates take only nucleotide substitutions into account. A comparable increase in variation in the African sample is not evident for insertion/deletion polymorphisms in the comparison with this European sample: 7 *vs.* 6 for *su(s)* and 16 *vs.* 11 for *su(w^a)*, respectively. These numbers of insertion/deletion variants are higher than those found previously (Aguadé *et al.* 1994) in a comparable sample from a North American population, 3 at *su(s)* and 6 at *su(w^a)*. The $\hat{\pi}^*$ estimates for both loci are smaller than those observed at X-linked loci in regions of normal crossing over in African (6.5×10^{-3}) and North American (3.5×10^{-3}) populations, as determined by averaging $\hat{\pi}$'s for *white*, *vermilion*, and *G-6-pdh* (Miyashita and Langley 1988; Begun and Aquadro 1993). The statistical significance of the reductions in polymorphism can be inferred under the assumptions of neutral theory (no linked selection, equilibrium between genetic drift and mutation to selectively equivalent alleles, and no recombination) and parameters,

TABLE 1

SSCP polymorphism at *su(s)* and *su(w^a)* in Africa

lines	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	**2**	1	1	1	1	1	1	1	1	1	*6*	2	9	1	1	*3*	1	1	1	1
2	8	1	1	1	1	6	1	1	1	3	3	2	*11*	1	1	*1*	1	*9*	1	1
3	**3**	1	1	**2**	1	1	1	1	**3**	**2**	1	1	**9**	5	3	*1*	2	1	1	1
4	3	1	1	1	1	1	**2**	**3**	**2**	1	*5*	2	1	1	1	*1*	**4**	1	1	1
5	**4**	1	1	1	1	1	1	1	3	1	*1*	3	*16*	5	3	*1*	1	1	1	1
6	3	1	1	**8**	**2**	**4**	1	1	1	1	**4**	**6**	*11*	1	1	*1*	1	2	1	1
7	**8**	1	1	**3**	1	**5**	**3**	1	1	*6*	3	2	**6**	1	1	*1*	**5**	1	1	1
8	3	1	1	9	1	1	2	1	**2**	1	1	2	1	5	3	*1*	1	1	1	1
9	3	1	1	1	1	6	1	1	2	1	1	1	9	5	3	*1*	1	1	1	**4**
10	**6**	1	1	1	1	**2**	1	1	4	1	**2**	**2**	6	1	**2**	*11*	1	4	1	**2**
11	8	1	1	**4**	1	6	1	1	1	3	1	2	1	1	1	*6*	2	1	1	1
12	3	1	1	1	1	**8**	1	1	1	1	1	**4**	1	*7*	**3**	*3*	1	1	1	1
13	3	1	1	**8**	1	1	1	1	1	1	*6*	2	1	1	1	*5*	2	1	1	1
14	3	1	1	9	1	1	2	1	2	1	1	2	**6**	1	2	*11*	1	4	1	2
15	3	1	1	1	1	1	1	1	**4**	**4**	1	1	*14*	1	1	*8*	1	1	1	1
16	*3*	1	1	1	1	2	1	1	4	1	*2*	**2**	1	1	*6*	*7*	1	**4**	1	**2**
17	**3**	1	1	1	1	**8**	1	1	4	1	1	1	6	**2**	1	*1*	1	4	1	2
18	8	1	1	1	*1*	6	1	**2**	1	1	**4**	2	**3**	6	1	*10*	1	1	1	1
19	1	1	1	**5**	1	1	1	1	4	1	1	1	**4**	**2**	1	*1*	1	4	1	2
20	8	1	1	1	1	6	1	2	1	1	4	2	**3**	**6**	1	*10*	1	1	1	1
21	8	1	**2**	4	1	**6**	1	1	1	3	1	*3*	1	5	3	*1*	1	1	1	*4*
22	3	1	1	9	1	1	2	1	2	1	1	2	1	1	1	*6*	**2**	**2**	1	1
23	**8**	1	1	1	1	6	1	1	1	3	3	2	*13*	1	1	*1*	1	*9*	1	1
24	3	1	1	**9**	1	1	2	1	2	1	1	2	6	1	2	*11*	1	4	1	6
25	3	1	1	**9**	*1*	1	**2**	1	2	1	1	2	6	1	1	*1*	1	4	1	**6**
26	3	1	1	1	1	8	1	1	1	1	1	**4**	**5**	**4**	3	*3*	1	*8*	1	1
27	*3*	1	1	1	1	8	1	1	1	**5**	**3**	2	**2**	1	1	*6*	1	1	1	6
28	3	1	1	1	1	1	*1*	1	1	1	1	1	1	5	3	*1*	2	1	1	1
29	3	1	1	1	1	**7**	1	1	1	3	1	2	*12*	**6**	2	*1*	1	**6**	1	1
30	8	1	1	1	1	**3**	1	1	1	3	1	2	**8**	1	1	*6*	1	**7**	1	1
31	*3*	1	1	1	1	1	1	1	**4**	**3**	1	2	6	1	1	*5*	1	1	1	1
32	*10*	1	1	1	1	1	1	1	4	1	1	1	*14*	1	1	*1*	1	1	1	1
33	**6**	1	1	1	1	2	1	1	4	1	4	2	6	1	*5*	*11*	1	4	1	6
34	3	1	1	1	1	1	1	1	4	1	1	1	**5**	1	*7*	*1*	1	2	1	1
35	**7**	1	1	1	1	1	1	1	2	1	1	1	**10**	1	1	*1*	1	**3**	1	1
36	3	1	1	1	1	8	1	1	1	1	**2**	1	9	**5**	*3*	*4*	1	1	1	1
37	6	1	1	1	1	**2**	1	**2**	1	1	**4**	2	*15*	1	1	*6*	1	1	1	**5**
38	3	1	1	1	1	1	1	1	2	1	1	1	**7**	1	1	*5*	1	1	1	1
39	1	1	1	**5**	1	1	1	1	4	1	1	1	*17*	*8*	1	*1*	1	4	1	**6**
40	3	1	1	**6**	1	1	1	1	4	1	*1*	1	1	**3**	3	*1*	**6**	1	1	1
41	8	1	**2**	4	1	6	1	1	1	3	*1*	**3**	1	5	3	*1*	1	1	1	1
42	3	1	1	1	1	1	1	1	2	1	1	1	6	1	**2**	*5*	1	**4**	1	1
43	3	1	1	**2**	1	1	*5*	1	**3**	**2**	1	1	**9**	**5**	**3**	*1*	**2**	1	1	1
44	8	1	2	**4**	1	6	1	1	1	3	1	**3**	1	5	3	*1*	1	1	1	1
45	**5**	1	1	1	1	1	1	1	1	3	1	**5**	*18*	1	1	*9*	2	1	1	1
46	3	1	1	1	1	1	1	1	4	1	1	1	1	1	**4**	*7*	1	*10*	1	6
47	3	1	1	1	1	1	1	1	1	3	1	2	1	5	3	*3*	1	1	1	1
48	3	1	1	1	1	1	1	1	**2**	1	*1*	1	1	1	1	*6*	2	1	1	1
49	8	1	1	**7**	1	**6**	1	1	*1*	**3**	**3**	2	1	1	1	*6*	**6**	1	1	1
50	**9**	1	1	5	1	1	1	1	1	1	1	1	*19*	1	1	*6*	2	**2**	1	1

SSCP classes for the various segments in each line. SSCP class numbers are along the top with the first and last base pair number proximal to the primer along the bottom. Boldface numbers in the table indicate segments that were sequenced. *Italicized* numbers indicate lines for which no SSCP classification was assigned. Each of these lines was directly sequenced and subsequently assigned to either a novel or existing SSCP class.

TABLE 2

su(s) and *su(w^a)* polymorphisms in Africa

	su(s)																																														
	1	1	1	1	1	1	1	1	1	1	1	3	3	3	4	4	4	4	4	4	4	4	5	6	6	6	6	6	7	7	7	7	8	9	9	9	9	10	10	10	10	11	11	11	11	12	12	12
lines	32	92	160	223	224	225-228	230	265	273	287	354	835	898	898+6	1084+	1106-1128	1186-1199	1203	1263	1303	1333	1397	1566	1793	1824	1831	1920	1942	2037	2039	2043	2063	2283	2413	2444	2448	2464	3629	3631	3680	3709	3937	3986	4042	4050	4104	4128	4187-4208
1	A	C	C	A	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	A	G	T	G	+
2	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	A	C	G	G	T	G	+
3	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	A	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	A	G	A	G	-	G	C	C	G	G	T	G	-
4	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	A	T	G	T	G	G	A	G	T	G	C	T	G	A	T	G	+
5	A	C	C	G	C	-	C	A	G	C	T	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	A	G	A	G	T	G	C	C	G	G	G	G	-
6	A	C	C	G	C	-	C	A	G	C	C	A	G	-	T	+	+	T	A	C	G	G	A	G	T	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	T	G	G	T	A	+
7	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	-	+	T	A	C	G	G	C	A	A	T	A	T	A	T	C	A	C	G	C	G	G	A	A	T	T	A	C	G	G	T	G	+
8	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	T	G	C	G	A	T	G	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	+
9	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	-
10	A	C	C	G	C	-	G	A	G	A	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	T	G	G	T	G	+
11	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	C	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	T	G	+
12	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	A	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	G	G	G	G	+
13	A	C	C	G	C	-	C	A	G	C	C	A	G	-	T	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	A	G	T	G	+
14	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	T	G	C	G	A	T	G	T	T	T	C	A	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	+
15	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	T	G	T	G	C	C	G	G	T	G	-
16	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	T	G	G	T	G	+
17	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	A	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
18	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	T	G	G	T	G	+
19	A	A	C	G	C	-	C	A	G	C	C	A	G	-	-	+	-	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
20	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	T	G	G	T	G	+
21	A	C	C	G	G	+	C	A	G	C	C	G	A	+	-	+	+	C	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	G	G	-
22	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	T	G	C	G	A	T	G	T	T	T	C	A	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	+
23	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	A	C	G	G	T	G	+
24	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	T	G	C	G	A	T	G	T	T	T	C	A	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	+
25	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	T	G	C	G	A	T	G	T	T	T	C	A	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	+
26	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	A	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	G	G	G	G	+
27	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	A	T	T	T	C	C	C	G	C	G	G	A	A	T	G	A	C	G	G	T	G	+
28	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	G	G	T	G	-
29	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	G	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	T	G	+
30	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	A	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	T	G	+
31	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	A	T	T	C	C	G	G	T	G	+
32	A	C	G	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
33	A	C	C	G	C	-	G	A	G	A	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	T	G	G	T	G	+
34	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
35	T	C	C	G	C	-	C	A	T	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	-
36	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	A	A	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	T	G	G	T	G	-
37	A	C	C	G	C	-	G	A	G	A	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	T	G	G	T	G	+
38	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	-
39	A	A	C	G	C	-	C	A	G	C	C	A	G	-	-	+	-	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
40	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	G	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
41	A	C	C	G	G	+	C	A	G	C	C	G	A	+	-	+	+	C	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	G	G	-
42	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	-
43	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	A	G	G	C	G	A	T	G	T	T	C	T	C	C	G	C	A	G	A	G	-	G	C	C	G	G	T	G	-
44	A	C	C	G	G	+	C	A	G	C	C	G	A	+	-	+	+	C	A	C	G	G	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	G	G	-
45	A	C	C	G	C	-	C	T	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	T	G	-
46	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	A	C	G	A	A	G	T	G	C	C	G	G	T	G	-
47	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	A	T	T	C	C	G	G	T	G	+
48	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	T	G	G	A	G	T	G	C	C	G	G	T	G	-
49	A	C	C	G	G	+	C	A	G	C	C	A	G	-	-	+	+	T	A	C	G	A	C	G	A	T	A	T	T	T	C	C	C	G	C	G	G	A	A	T	T	A	C	G	G	T	G	+
50	A	C	C	G	C	-	C	A	G	C	C	A	G	-	-	+	-	T	A	C	G	G	C	G	A	T	G	T	T	T	C	C	C	G	C	G	G	A	G	T	G	C	C	G	G	T	G	-

(*continued*)

TABLE 2

(Continued)

	su(w^a)																																																											
	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	14	14	14	14	15	15	15	15	16	16	16	16	16	16	16	16	16	16	17	17	17	17	18	18	18	18	18	18	18	18	18	18	20	20	20
lines	60	98	101	109	113+	117+	122	122+2.1	122+2.2	125	127+	130	131-136	137	138	142	153-155	156	161	165+12.1-.2	165+12.4	165+12.6	165+12.9	165+12.10	179	210	251	296	330-331	421	457	891	893-894	928	1020	1100	1121	1137	1138	1195	1196	1197	1223+6	1232	1254	1523	1526	1539	1644	1941	1942	1957	1977	1981+	2020	2067	2099	2115	2131	2550	2579	2589
1	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	G	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
2	T	A	T	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	G	T	G	C	T
3	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
4	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	G	C	C	A	T	A	-	T	C	G	T	T	G	C	T
5	T	A	G	A	-	C	T	-	-	A	A	A	+	G	A	T	+	G	A	+	T	T	A	C	T	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
6	T	A	T	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	A	C	G	T	T	G	C	T
7	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	T	G	T	T	C	A	T	A	-	T	C	G	T	T	G	C	T
8	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
9	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	C
10	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	+	A	A	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
11	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
12	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	T	C	+	C	G	C	G	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
13	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	+	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
14	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	+	A	A	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
15	T	T	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	T	G	A	T	A	G	C	A	+	C	C	C	-	C	G	A	A	G	G	G	A	A	+	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
16	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	+	T	G	C	A	G	G	T	T	-	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
17	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	G	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
18	C	T	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	C	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	-	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
19	T	T	G	A	-	-	T	-	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	G	C	G	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
20	C	T	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	C	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	-	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
21	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	C
22	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	A	C	G	T	T	G	C	T
23	T	A	T	A	-	-	T	-	-	A	A	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	G	T	G	C	T
24	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	+	A	A	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
25	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
26	T	A	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	T	G	A	C	A	G	C	A	+	T	C	C	+	C	G	C	G	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	G	G	T	C	G	T	T	G	C	T
27	T	A	G	A	T	-	T	T	T	G	A	A	+	G	A	A	+	G	A	+	T	T	A	T	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	A	C	T
28	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
29	T	A	G	A	-	-	T	-	-	A	-	A	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	C	A	-	T	C	A	T	T	G	C	T
30	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	-	T	T	A	T	A	A	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	T	A	T	A	-	T	C	G	T	T	G	C	T
31	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	+	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
32	T	T	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	T	G	A	T	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
33	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	+	A	A	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
34	T	A	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	A	+	T	G	A	C	A	G	C	A	+	C	C	T	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	A	C	G	T	T	G	C	T
35	T	A	G	A	-	-	T	-	-	A	A	A	+	G	A	A	+	G	T	+	T	G	A	T	A	G	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	A	G	T	A	G	C	T
36	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
37	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	+	G	A	+	T	T	G	T	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	A	T
38	T	T	G	G	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	G	C	A	G	G	G	A	A	+	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
39	T	T	G	A	-	-	T	-	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	G	+	C	C	C	-	C	G	C	A	C	G	G	A	A	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
40	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	A	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
41	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
42	T	T	G	A	-	-	T	T	-	G	A	A	-	A	T	A	-	G	A	+	T	T	A	T	A	A	C	A	+	C	C	C	-	C	T	C	A	G	G	G	A	A	+	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	G	C	T
43	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
44	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	A	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
45	T	A	G	A	-	-	-	-	-	A	A	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	+	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
46	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	+	T	G	C	A	G	G	T	T	-	-	A	C	C	G	T	C	C	G	T	A	-	T	C	G	T	T	A	C	T
47	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	-	C	C	C	+	C	G	C	G	G	G	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
48	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
49	T	A	G	A	-	-	T	-	-	A	-	G	+	G	A	A	+	T	A	+	T	T	A	C	A	G	C	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	A	T	C	C	A	T	A	-	T	C	G	T	T	G	C	T
50	T	T	G	A	-	-	T	-	-	A	A	A	-	A	T	A	-	G	A	-	T	T	A	T	A	A	-	A	+	C	C	C	-	C	G	C	A	G	A	G	A	A	-	A	C	C	G	T	C	C	A	T	A	-	A	C	G	T	T	G	C	T

(*continued*)

$3N\mu = \theta$ equal to average $\hat{\theta}$'s observed at *white* and *vermilion* and *G-6-pdh* in African and non-African samples: 0.0096 and 0.0038, respectively (Miyashita and Langley 1988; Hudson 1990; Begun and Aquadro 1993). The probabilities of the observed or fewer segregating nucleotide sites (Hudson 1990) are as follows: *su(s)* African, ≤ 0.001; *su(s)* Europe, ≤ 0.001; *su(s)* North America, ≤ 0.001; $su(w^a)$ African, <0.03; $su(w^a)$ Europe, <0.08; and $su(w^a)$ North America, <0.04.

To examine the site frequency spectra, Table 5 presents the calculated Tajima's *D* (Tajima 1989) for each locus population. Under an equilibrium between random genetic drift and selectively neutral mutation, the expected value of Tajima's *D* in samples such as these is close to zero. A skew in the frequency spectrum (excess rare polymorphisms) yields a negative value for Tajima's *D*. The distribution of Tajima's *D* is simulated as in Braverman *et al.* (1995) for 10,000 replicas (for the actual sample sizes and the observed number of segregating sites, assuming no intragenic recombination). The proportion of those simulated samples with Tajima's *D* values less than the observed, $P\ \{\text{Tajima's } D \leq D_0\}$, is an estimate of the probability of the observed value of D_i or less given S_i segregating sites. The observed *D* values for *su(s)* from the European sample are significantly negative ($P < 0.05$) compared to the expectations of the neutral theory and the background selection model, while the critical values associated with that for $su(w^a)$ are 0.07. Tajima's *D* values from the African sample are also negative, but not statistically significant. Wall (1999) has demonstrated that the assumption of no intragenic recombination can lead to excessively conservative tests of Tajima's *D* (see also Braverman *et al.* 1995). Simulations assuming a more realistic level of recombination ($3Nc = 5$) yield less conservative critical values (see below).

The linkage disequilibrium between sites is depicted in Figures 1a and 2a, which also show the noncoding regions (3220 bp) of the *su(s)* locus surveyed by SSCP analysis of the indicated fragments for the African and European samples. The ticks on the horizontal line below the SSCP fragments represent the positions of informative sites (at least two of the rarer state observed in the sample). The shading of the circles connected to the ticks by the dashed lines increases with the expected heterozygosity at the site.

The diagonal matrix at the bottom presents the statistical significance of the nonrandom associations between the pairs of sites. As expected, closely linked sites with higher expected heterozygosities are more often represented among the pairs showing statistically significant nonrandom association. More than 10% of all pairs of sites at *su(s)* showed nonrandom associations with $P < 0.005$ in the African sample (Figure 1a). In the European sample (Figure 2a) 4 of 28 comparisons were significant at the $P < 0.01$ level. Also shown (small solid triangles) are the positions of the "minimum number" exchanges in the history of the sample (Hudson and Kaplan 1985): six in the African sample and three in the European sample.

Figures 1b and 2b present the same types of results for the surveys of several noncoding regions of the $su(w^a)$ locus in the African and European samples, respectively. A small block of many highly polymorphic sites between positions 98 and 156 exhibits strong linkage disequilibrium in the sample from Africa (Figure 1b), although

TABLE 2

(Continued)

Distribution of DNA sequences in 12 segments of the *su(s)* region and 8 segments of the $su(w^a)$ region for 50 *X* chromosome lines from a natural population of *D. melanogaster* from Zimbabwe, Africa. Observed (boldface) and inferred sequence polymorphisms among the 50 lines. The base numbering (top row in table) is consistent with that of the published sequence (GenBank accession nos. M57889 and X06589). Minus signs in the table represent either deletions or nonpresence of insertions. Single base insertions are indicated by naming the inserted base and are given the position of the previous base number followed by a plus sign (for example, in column 1084+ in the *su(s)* region, there is a T residue inserted in lines 6 and 13 after the base number 1084). Multiple base insertions are scored simply with a plus sign indicating presence if the insertion is monomorphic in sequence for all alleles containing it. They are numbered with the number of the previous base followed by a plus sign and the size of the insertion (for example, in column 898+6 in the *su(s)* region, there is a 6-bp insertion (TATATA) following base number 898 in lines 21, 41, and 44). $su(w^a)$ 1223+6 is ATTCAT. There were two instances where insertions were not monomorphic in sequence. These were an insertion after base number 122 (TT or T) and a 12-base insertion after base 165 in the $su(w^a)$ region (GTGTCTCAATTT). In both of these cases, the polymorphic bases are listed as decimals after the insertion size number (for example, in 165+12.4, lines 18 and 20 have a C residue at the fourth site in the 12-bp insertion where all of the rest of the lines have a T). All of the Zimbabwe lines carried some form of the 12-bp insertion after base 165 in $su(w^a)$, but this is not true of other populations surveyed. There are 9 sites where there is variation within an SSCP class for a polymorphism. At *su(s)* base 2241, lines 18 and 37 (SSCP 2) differed for an A to G substitution, which was present in 37 but not in 18. These same lines differed again at *su(s)* 2277, where 18 had an A to T substitution not shared by 37. At $su(w^a)$ base 214, lines 12 and 49 (SSCP 1) differed for a C to G substitution, with 12 showing presence of the substitution. At $su(w^a)$ base 478, line 29 has a G to A substitution that line 20 does not, creating a discrepancy in SSCP 6. Again in these lines, at $su(w^a)$ base 509, 20 has a C to T substitution that 29 does not. At $su(w^a)$ base 507, lines 36 and 43 (SSCP 5) differed for a C to T substitution, which only 36 carried. At $su(w^a)$ base 1486, lines 43 and 22 (SSCP 2) differ for a G to A substitution, which 43 carries. These lines also differ at $su(w^a)$ base 1512, where 22 has a C to T substitution that 43 lacks. At $su(w^a)$ base 1989, line 42 differs from lines 16 and 46 (SSCP 4), where 42 lacks a T to C substitution that the other two contain. All 9 of these sites were dropped from the survey.

at least two exchanges are inferred to have occurred in the ancestry of these sampled alleles within this small fragment. The pattern in the remaining portion of the matrix of nonrandom associations appears similar to that for *su(s)* from Africa, 24 of 276 pairs significant with $P < 0.005$. The distribution of these significant comparisons shows little apparent association with the distance between pairs (discussed below). There are a minimum of 10 exchanges in the history of these alleles. The European sample of $su(w^a)$ alleles does not exhibit the high level of polymorphism in the first fragment (number 13). The most significant associations are clustered among the tightly linked pairs of sites. A total of 21 of the 153 comparisons are significant with $P < 0.005$; 9 of these 21 are between adjacent (tightly linked) sites. At least 5 exchanges among the ancestral alleles of those in the European sample can be deduced.

The distribution of linkage disequilibria within and between *su(s)* and $su(w^a)$ is similar to that observed in the survey of the North American population (Aguadé *et al.* 1994). The estimate of crossing over between $su(w^a)$ and *su(s)* is less than 10^{-3} (see below). As was seen in the North American sample, linkage disequilibria between sites at the two loci are rare. In the European sample none of the 144 pairs of polymorphic sites, 1 at *su(s)* and the other at $su(w^a)$, are in linkage disequilibrium; *i.e.*, $P < 0.05$. Of the 1008 such pairs in the African sample 80 appear to be in disequilibrium: 7 with $P < 0.005$, 2 with $P < 0.001$, and the remainder with $P < 0.05$; the mean squared correlation coefficient $r^2 = 0.035$ (see discussion).

DISCUSSION

This survey brings two new aspects to the study of the reduction in DNA sequence polymorphism in regions of reduced crossing over per physical length in natural populations of *D. melanogaster*. First, two additional populations (African and European) are surveyed for *su(s)* and $su(w^a)$, providing a more complete view of the pattern of DNA sequence polymorphism at $su(w^a)$ and *su(s)* in *D. melanogaster* throughout its distribution. Second, the increased overall levels of DNA sequence polymorphism in the African population provide more accurate estimates and more powerful statistical inferences than were available in previous surveys of regions of low crossing over. Africa is thought to be the ancestral home of *D. melanogaster* (David and Capy 1988). It is not known when this species spread out of Africa to become the cosmopolitan species it is now. It may well have accompanied human ancestors as they migrated out of Africa. In any case, much like their human commensals (Cavalli-Sforza *et al.* 1994), African *D. melanogaster* harbors most of the DNA sequence polymorphism found throughout the rest of the world and is significantly more heterozygous than populations on other continents (Begun and Aquadro 1993). It seems probable that African populations have been prehistorically larger and ecologically more stable than populations on other continents that were recently established. African populations are likely to be closer to the idealized populations assumed in theoretical analyses, while non-African populations may still be undergoing more intense adaptive selection to the non-African environments. For these statistical and biogeographical reasons the African sample affords the most straightforward interpretation.

The *su(s)* and $su(w^a)$ region exhibits reduced crossing over per physical length. The map distance to *y* of genes near this *X* telomere is the most appropriate available measure of the crossing over per physical length for two reasons. First, it is the most direct and reliable quantitative observation. Second, and more important, is the asymmetry in the pattern of crossing over per physical length in this genomic region. That portion of the genome that is tightly linked to *su(s)* and $su(w^a)$ is toward *yellow* and extends out to the *X* telomere. Crossing over between centromere proximal markers and *y* continues to decline distally throughout cytological section 1. Crossing over between *y* (cytological position 1B1; map position 0.0) and $su(w^a)$ (1E1-4) is reported to be somewhat $<10^{-3}$ (M. Green, personal communication). R. Voelker and J. Mason (personal communication) report that rare recombinants between *y* and recessive lethals adjacent to *su(s)* (1B13) occur with a frequency somewhat $<10^{-4}$. And since meiotic crossing over near *y* and beyond is absent, this terminal region forms a "block" of loci, all at the same (crossing over) distance from *su(s)* or $su(w^a)$. Crossing over is increasing so rapidly in the centromere-proximal direction that the impact of linked selected variation must be much less. For example, the *white* locus (3C2) crosses over with *y* ~3% of the time.

su(s) and $su(w^a)$ are thus tightly linked to ~1% of the genome (cytological section 1) that undergoes little crossing over. This large reduction in crossing over per physical length can be compared to the reductions in polymorphism at the loci in this region. The average number of pairwise differences per site (and $\hat{\theta} = 3N\mu$) of loci in *X* chromosome regions of normal crossing over (*white, vermilion,* and *G-6-pdh*) is 0.007 (0.01) for Africa and 0.004 (0.004) for North America (Miyashita and Langley 1988; Begun and Aquadro 1993). The estimates $\hat{\pi}^*$ for $su(w^a)$ are half, while those at *su(s)* are about one-fourth (see Table 5), consistent both with the linkage to *y* and with average levels of variation across populations. $\hat{\pi}$ for the *y-ASC* is reported to be 0.001 in both African and non-African populations (Aguadé *et al.* 1989; Martín-Campos *et al.* 1992; Begun and Aquadro 1993), which fits well into this pattern. Such simple comparisons of the levels of polymorphism will not discriminate among the proposed linked-selected-perturbation models. Any quantitative interpretation of a two- or fourfold reduction in expected heterozygosity depends on independent knowledge of several addi-

TABLE 3

SSCP polymorphism at *su(s)* and *su(w^a)* in Europe

lines	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	1	1	1	1	1	1	1	1	**3**	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	**4**	1	1	1	2	1	1	1	1	1	1
3	1	3	1	*1*	1	1	1	1	1	1	1	1	1	**4**	1	1	1	1	1	1
4	1	3	1	***3***	1	1	1	1	1	1	1	1	1	**6**	1	1	1	1	1	1
5	1	**2**	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	**3**	1	1	1	**2**	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	**3**	1	1	**2**	**2**	1	*1*	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	2	3	1	1	1	1	1	1	1	1	**2**	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1	**3**	1	1	1	1	1	1	1	1	1	1	1	**2**	2	1	1	1	1	1
15	2	1	1	1	1	1	1	1	1	1	**2**	1	1	1	1	1	1	2	1	1
16	1	1	1	**2**	1	1	1	1	1	1	1	1	1	1	**5**	**4**	1	1	1	1
17	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	**2**	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1	1	1	**2**	**3**	**2**	1	1	**2**	1	1
20	**2**	1	1	1	1	1	1	1	1	1	1	1	1	1	5	1	*1*	1	1	1
21	1	1	1	**2**	1	1	1	1	1	1	1	1	**2**	**3**	2	1	1	**2**	1	1
22	2	3	1	1	1	1	1	1	1	**3**	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	5	**4**	1	1	1	1
24	**2**	**3**	1	1	1	**3**	1	1	1	1	1	1	**3**	**7**	1	**3**	*1*	1	1	1
25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	*6*	1	1	1	1	1
27	2	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**5**	*4*	1	1	1	1
29	**2**	1	1	1	1	1	1	1	1	**3**	*1*	1	1	1	1	1	1	1	*1*	1
30	2	1	1	1	1	1	1	1	1	3	*5*	1	1	1	1	1	1	1	1	1
31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
32	1	1	1	1	1	1	1	1	1	**2**	1	1	**5**	5	*5*	2	1	1	1	1
33	1	1	1	*1*	1	1	1	1	1	1	1	1	1	1	2	1	1	1	**2**	1
34	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
36	1	1	1	1	1	1	1	1	1	1	1	1	**4**	**5**	**3**	**2**	1	1	1	1
37	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	*1*	1	1	1
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
41	1	1	1	1	1	1	1	1	1	1	**3**	1	1	1	2	1	1	1	1	1
42	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	**2**	1
43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	*7*	1	1	1	1	1
44	1	1	1	1	1	1	1	1	1	1	*4*	1	**4**	**5**	**3**	**2**	1	1	1	1
45	1	1	1	1	1	**2**	1	1	1	1	1	1	1	1	2	1	1	1	1	1
47	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
49	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
52	1	1	1	1	1	1	1	1	1	1	1	1	1	*2*	2	1	1	1	1	1

SSCP class assignments for each fragment in each line. Boldface numbers indicate lines for which each fragment was sequenced. *Italicized* numbers indicate lines for which no SSCP classification was assigned. Each of these lines was directly sequenced and subsequently assigned to either a novel or existing SSCP class.

tional parameters (beyond crossing over per physical length) in all the proposed models. For directional hitchhiking the overall reduction depends on both the selection coefficient(s) and on the rate per centimorgan per generation at which rare favored alleles fix (Kaplan *et al.* 1989). Similarly the background selection model can predict such a reduction in the expected heterozygosity under a wide range of selection coefficients and deleterious mutation rates (Hudson and Kaplan 1995; Charlesworth 1996). In the random-environment-hitchhiking models that Gillespie (1997) analyzed the parameters are also difficult to independently ascertain. To discriminate among these models other properties of the polymorphism offer better prospects.

The frequency spectra at loci demonstrating clear reductions in expected heterozygosity associated with reduced crossing over per physical length offer some hope. Braverman *et al.* (1995) showed that simple directional hitchhiking, which reduces heterozygosity by these observed amounts, should lead to a strong skew in the frequency spectra. This distortion from expectation under selective neutrality should be detectable in samples of the size reported here. Background selection is not expected to produce detectable skew (Charlesworth 1996). While Gillespie's analyses of various random environment models do not provide sample properties nor any indication of statistical power, they do show a clear correlation between the reduction in average heterozygosity and the skew toward an excess of rare polymorphisms (Gillespie 1997). As Table 5 shows, the frequency spectra of nucleotide polymorphisms at both *su(w^a)* and *su(s)* in both the African and European samples are skewed toward rare sites unlike those observed previously in the North American sample (Aguadé *et al.* 1994). The European samples reach nominal statistical significance in the deviations of Tajima's *D* from the predictions of the neutral model or background selection. The results from the survey of *su(s)* and *su(w^a)* from Europe and Africa favor a directional hitchhiking explanation over background selection. But how well do the observations fit the predictions of directional hitchhiking? Using the simulation approach in Braverman *et al.* (1995), the rates of directional hitchhiking (and selection coefficient) were set to yield the observed reductions in average number of pairwise differences per site from those expected in regions of normal crossing over, 0.007 in Africa and 0.004 in Europe and North America. Figure 3 shows *above* the probabilities of the observed Tajima's *D*'s or less under the neutral or background selection models for the six locus samples. Depicted *below* the *x*-axis in Figure 3 are the proportions of directional hitchhiking simulations in which the Tajima's *D* value exceeded that observed. The lack of linkage disequilibrium between these two loci suggests that the similarities are unlikely due to a single recent event.

It is evident in Figure 3 that the North American sample (especially the *su(s)* result) appears exceptional. Since the African, European, and North American populations cannot be considered strictly independent, it is prudent to take the African results as the most representative, since they are based on many more segregating sites and are from the putative ancestral region. While neither the background selection and neutral models nor the directional hitchhiking model can be rejected by Tajima's *D* values from the African sample, these results favor some form of hitchhiking. And, indeed, the European results also support hitchhiking over background selection. The inconsistency of the North American results with this pattern might be attributed to transient hitchhiking associated with the more recent colonization of the Western Hemisphere (David and Capy 1988).

It has been suggested from empirical studies (Stephan and Langley 1989; Begun and Aquadro 1993) and from theoretical modeling (Nordborg *et al.* 1996; Stephan *et al.* 1998) that geographic differentiation in DNA sequence polymorphism might be greater in those

TABLE 4

su(s) and *s(w^a)* polymorphisms in Europe

	su(s)														*su(w^a*																														
	1	1	2	4	4	4	6	6	10	10	10	11	11	11	13	13	13	13	13	13	13	13	13	13	13	13	13	14	14	14	14	15	15	15	15	15	16	16	16	16	18	18	18	19	19
lines	224	225-228	698-703	1049	1324	1410	1841	1920	3560-3562	3586	3667	3888	3898-3927	3937	113	122+2	125	127+	130	131-136	137	138	153-155	156	165+12*	183	210	330-331	478	483	507	889	891	893-894	898	1007-1021	1049-50	1054	1121	1138	1957	1983-2005	2099	2199	2201
1	G	+	+	T	A	A	C	G	+	C	A	G	-	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
2	G	+	+	T	A	A	C	G	+	C	T	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	A	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
3	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
4	G	+	+	-	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	C	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
5	G	+	-	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
6	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
7	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
8	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
9	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
10	G	+	+	T	A	A	C	G	+	C	A	G	-	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	A	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
11	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
12	C	-	+	T	A	A	C	G	+	C	A	-	+	C	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
13	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
14	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	A	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
15	C	-	+	T	A	A	C	G	+	C	A	-	+	C	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	C	-	A	T	G
16	G	+	+	T	G	T	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	C	+	A	+	-	T	A	G	T	+	G	T	G
17	C	-	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
18	G	+	+	T	A	A	C	G	+	T	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
19	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	A	+	G	A	+	T	T	T	G	+	A	A	C	A	T	-	A	+	+	A	A	G	C	-	A	T	G
20	C	-	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	C	+	A	+	+	A	A	G	T	+	G	T	G
21	G	+	+	T	G	T	C	G	+	C	A	G	+	A	G	-	A	-	A	+	G	A	+	T	T	T	G	+	A	A	C	A	T	-	A	+	+	A	A	G	C	-	A	T	G
22	C	-	+	T	A	A	C	G	-	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
23	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	C	+	A	+	-	T	A	G	T	+	G	T	G
24	C	-	+	T	A	A	G	A	+	C	A	G	+	A	-	+	G	A	A	-	A	T	-	G	T	G	A	+	G	A	T	C	C	+	A	+	+	A	G	G	T	+	G	T	G
25	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
26	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	T	-	A	+	+	A	A	G	T	+	G	T	G
27	C	-	+	T	A	A	C	G	-	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
28	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	C	+	A	+	-	T	A	G	T	+	G	T	G
29	C	-	+	T	A	A	C	G	-	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
30	C	-	+	T	A	A	C	G	-	C	A	-	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
31	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
32	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	G	-	T	G	+	G	A	C	A	C	+	A	+	+	A	A	A	T	+	G	T	G
33	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	-	C
34	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
35	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
36	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	G	-	T	G	+	G	A	C	A	C	+	C	-	+	A	A	A	T	+	G	T	G
37	C	-	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
38	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
39	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
40	C	-	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
41	G	+	+	T	A	A	C	G	+	C	A	G	-	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
42	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	-	C
43	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	T	-	A	+	+	A	A	G	T	+	G	T	G
44	G	+	+	T	A	A	C	G	+	C	A	G	+	C	G	-	A	-	G	+	G	A	+	G	-	T	G	+	G	A	C	A	C	+	C	-	+	A	A	A	T	+	G	T	G
45	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G
47	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
48	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
49	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
50	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
51	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	G	A	C	C	C	+	A	+	+	A	A	G	T	+	G	T	G
52	G	+	+	T	A	A	C	G	+	C	A	G	+	A	G	-	A	-	G	+	G	A	+	T	C	T	G	-	A	A	C	A	T	-	A	+	+	A	A	G	T	+	G	T	G

(*continued*)

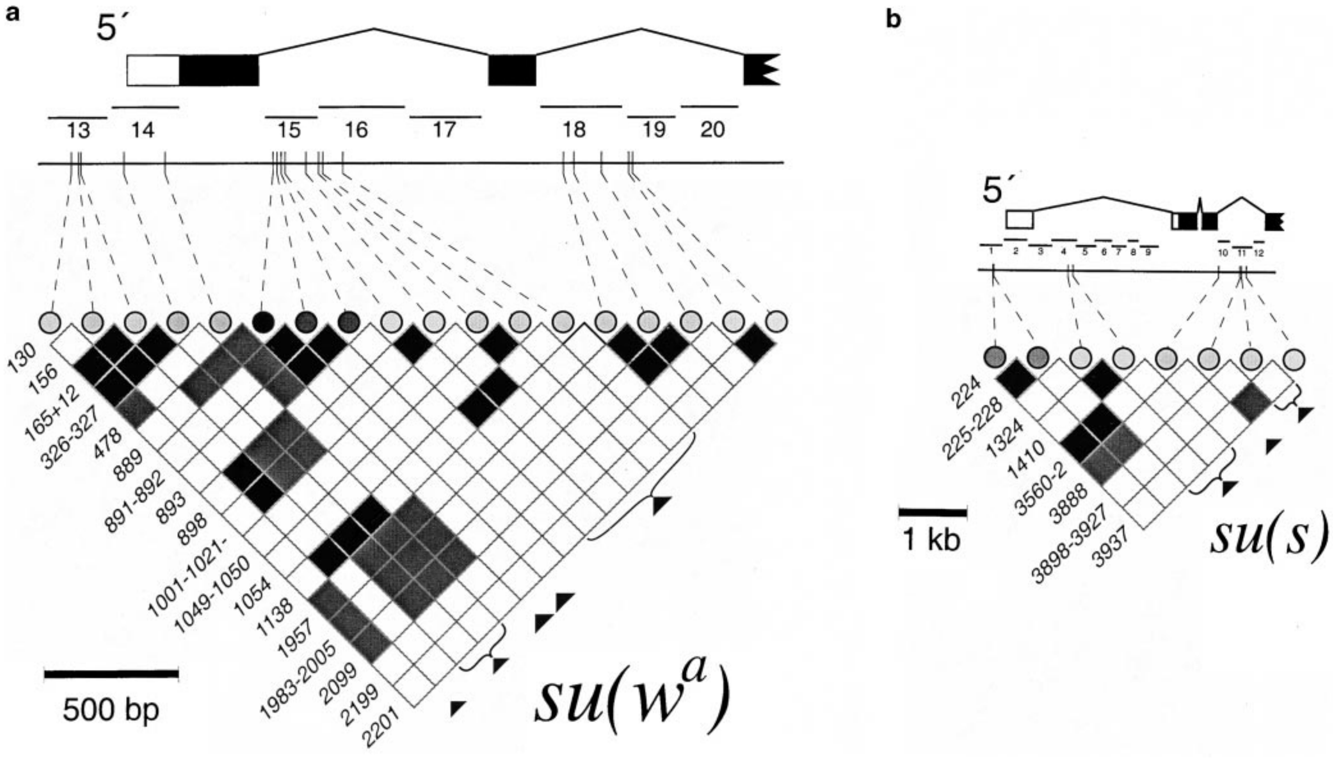

Figure 2.—Polymorphism and linkage disequilibria in the European sample. The positions of polymorphic sites detected (ticks on the third level) in the survey of the 5′ portion of the *su(w^a)* (a) and of *su(s)* (b) in 51 alleles from a European population are indicated below a depiction of the gene structure (the open box is the 5′ untranslated portion, the solid boxes are coding portions of exons, and the thin lines are the introns). Only those polymorphisms at which the least common state occurred twice or more in the sample are shown. Directly below the gene structure are the positions and sizes of the PCR fragments surveyed by SSCP and DNA sequencing. The triangular matrix of squares represents the statistical significance determined by Fisher's exact test (uncorrected for multiple tests); white, $P > 0.05$; light gray, $P < 0.05$; dark gray, $P < 0.01$; and black, $P < 0.005$. The dashed lines connect the polymorphic sites to their corresponding columns in the matrix. The shading of the circles at the top of each column increases with the expected heterozygosity of the polymorphic site. Along the left margin of the matrix are the positions in the published sequence of the corresponding polymorphic sites. Along the right margin are the inferred positions of the "minimum" number of recombination events in the history of the sample (Hudson and Kaplan 1985).

regions of the genome where crossing over per physical length is reduced. But the inherent paucity of variation and other considerations place considerable limitations on empirically based conclusions (Charlesworth 1998). The consideration of the genetic differentiation among populations of *D. melanogaster* has two broad aspects: the relatively greater overall level of polymorphism in Africa and the variance in site frequencies among populations. Figure 4, a–d, shows the distributions of F_{ST} of individual sites for *su(s)*, Europe and Africa; *su(s)*, North America and Africa; *su(w^a)*, Europe and Africa; and *su(w^a)*, North America and Africa, re-

TABLE 4

(Continued)

Distribution of DNA sequences in 12 fragments in the *su(s)* region and 8 fragments in the *su(w^a)* region for 51 *X* chromosome lines derived from a natural European population of *D. melanogaster*. Observed (boldface) and inferred sequence state at all polymorphic sites among the 51 lines. The base numbering (top row in table) is consistent with that of the published sequence (GenBank accession nos. M57889 and X06589). As in Table 1, minus signs indicate a deletion of the specified bases or the nonpresence of an insertion. Single base insertions are indicated by naming the inserted base and are given the position of the preceding base followed by a plus sign. Monomorphic multiple base insertions are indicated by a plus sign and are numbered with the position of the preceding base followed by a plus sign and the number of inserted bases. In *su(w^a)*, a 12-base insertion at position 165 (marked with an asterisk) was not monomorphic in the alleles containing it. In this case, lines that contain no form of this insertion are indicated with a minus sign. Lines that contain the inserted sequence, gtgtctcaaTtt, are labeled with a T. Lines that contain the inserted sequence, gtgtctcaaCtt, are labeled with a C. There was one instance of sequence variation within an SSCP class in the European population. In *su(s)* fragment 10, lines 18 and 32 (SSCP 2) were either T or C, respectively, at position 3586. However, these were the only two lines assigned to SSCP class 2 in this fragment. There were no lines in which the nucleotide state at that position was ambiguous so the site was retained in the analyses.

TABLE 5

Summary statistics

				S	Tajima's D		
Population	Locus	$\hat{\pi}^*$	Bootstrap	nt (i/d)	nt	i/d	Pooled
Africa	*su(s)*	1.82×10^{-3}	1.70–2.05	41 (7)	−1.28	−0.47	−1.19
Europe	*su(s)*	0.35×10^{-3}	0.24–0.44	8 (6)	−1.54*	−1.02	−1.50*
North America	*su(s)*	1.02×10^{-3}	0.93–1.09	10 (3)	+1.31	−0.14	+1.01
Africa	*su(w^a)*	4.67×10^{-3}	4.28–5.19	46 (16)	−1.04	−0.17	−0.83
Europe	*su(w^a)*	1.25×10^{-3}	1.02–1.83	20 (11)	−1.38	−1.47	−1.53*
North America	*su(w^a)*	2.08×10^{-3}	1.99–2.36	17 (6)	−0.40	+0.21	−0.23

Estimates of the average and of the 95% bootstrap confidence intervals of the number of pairwise differences per site; the observed number of nucleotide site polymorphisms (S) among the sampled alleles; and Tajima's D for nucleotide (nt) site substitutions, insertion/deletion variants (i/d), and for both pooled. The results from Aguadé *et al.* (1994) for a North American sample are included for comparison. * indicates statistical significance, *i.e.*, $P\{\text{Tajima's } D \leq D_0\} < 0.05$ (see text).

spectively. The patterns of these distributions are remarkably consistent in both their shape and the large contribution of polymorphism of the African sample. There is no evidence of "fixed differences" between populations. The mean F_{ST} values (Hudson *et al.* 1992) for *su(s)* and *su(w^a)* between the African and European (and North American) samples are 0.153 (0.245) and 0.291 (0.343), respectively. Comparable values for loci in regions of normal crossing over have been reported (*white*, 0.28; *vermilion*, 0.32; *G-6-pdh*, 0.30; *Pgd*, 0.25), while greater differentiation was reported for three loci in regions of low crossing over per physical length [*yellow*, 0.56; *achaete*, 0.54; *su(f)*, 0.60; Begun and Aquadro 1993]. This difference in the apparent level of geographic differentiation can be attributed to the small number of polymorphic sites surveyed at the three loci in that study.

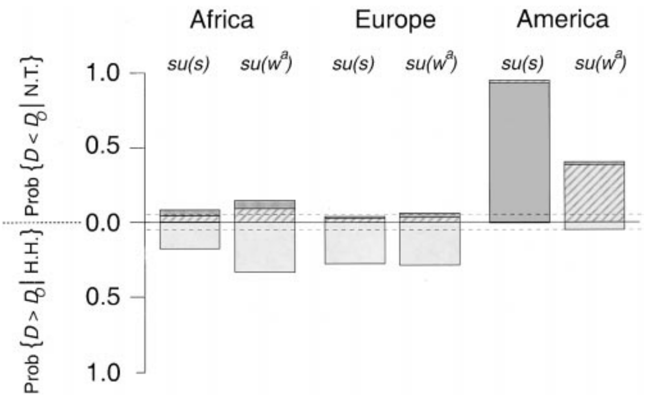

Figure 3.—The estimated relative probabilities of Tajima's D values less than that observed under the neutral or background selection models, "N.T.," or greater than that observed under the directional hitchhiking model, "H.H." Neutral coalescent simulations (Hudson 1990) were conducted to generate 10,000 samples of size 50 for each locus in each population. Given the observed numbers of segregating sites the distribution of Tajima's D was obtained. The proportion of simulations yielding a value less than that observed in the data (Table 3) is plotted above (gray bars for $3Nc = 0.0$; hatched bars for $3Nc = 5.0$). Simulations of recurrent, random directional hitchhiking (Braverman *et al.* 1995) were conducted with the hitchhiking intensity sufficient to reduce the average number of pairwise differences per site to the observed $\hat{\pi}^*$ values from a neutral expectation of 0.007 for Africa and to 0.004 for Europe and North America. The proportion of 10,000 such simulations yielding a Tajima's D value greater than the observed is plotted below (open bars). The horizontal dashed lines indicate $P > 0.05$.

Figure 5 shows the distributions of the linkage disequilibria (measured as squared correlation coefficient, r^2) between pairs of sites plotted against distances (base pairs) for *su(w^a)* and *su(s)*, respectively, for the African sample (the sample with the most polymorphism and thus the most linkage disequilibrium information). Also, as argued above from a biogeographic perspective, the African population may be closer to equilibrium with respect to mutation, drift, selection, and recombination. Only sites where the rarer state occurred at least twice in the sample can be considered in the linkage disequilibrium analysis. Clearly most of the large linkage disequilibria ($r^2 > 0.5$) occur between sites separated by small distances, <200 bp. Also shown are the means over six contiguous intervals of distance [63 and 105 r^2 values averaged in each interval for *su(s)* and *su(w^a)*, respectively]. As expected from previous surveys and theoretical predictions, the scatter is large for individual r^2 values. The distribution of the mean r^2 values reinforces the view that linkage disequilibrium tends to dissipate quickly and is near that expected from sampling, 0.02 for distances >500 bp.

To summarize the distributions of r^2 an empirical function was developed and fitted. Under the assumption of $r = 0.0$, the expected value of r^2 is the reciprocal of the sample size, 0.02 in this case. Under Wright-Fisher sampling the mean value of r^2 in the absence of any recombination can be estimated by simulation (Hudson 1990). Specifically an estimate of this "intercept" was derived from 1000 neutral, infinite site model coa-

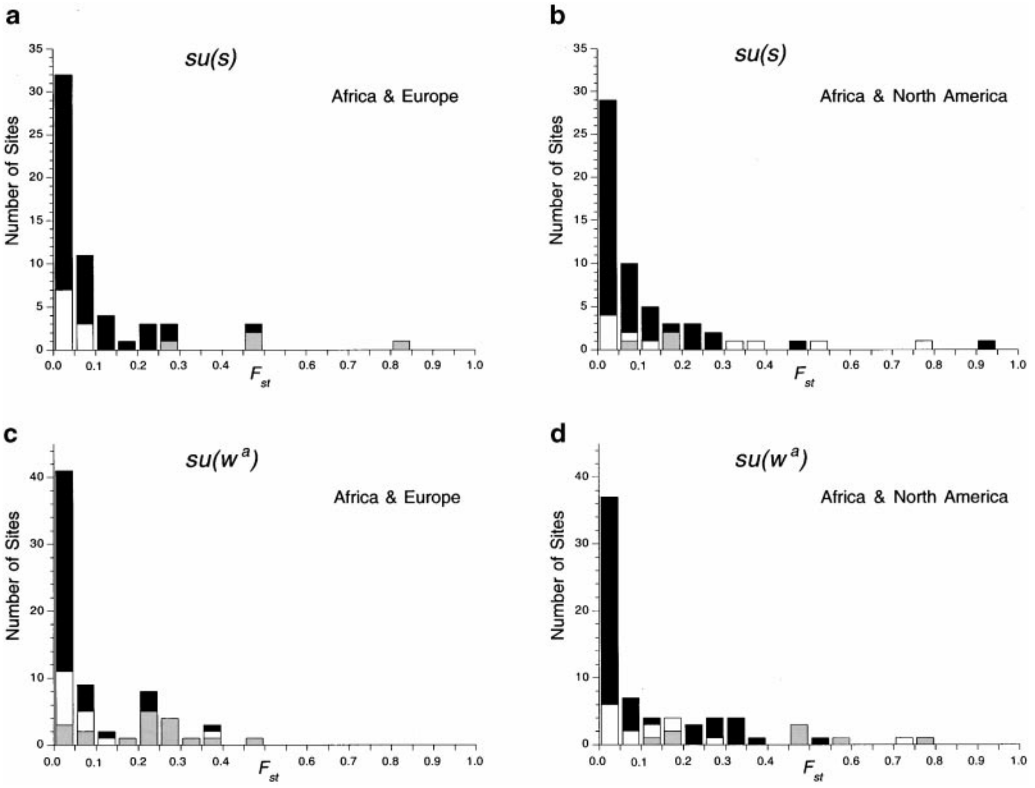

Figure 4.—Distributions of F_{ST} values for individual polymorphic sites at the *su(s)* and *su(w^a)* loci in comparisons of the African sample with the European (a and c) and the North American (b and d). The F_{ST} was calculated as $1 - (H_w/H_b)$, where H_w is the unweighted average expected heterozygosity and H_b is the expected heterozygosity of an unweighted pooled population. The solid portion of each bar reflects the number of sites polymorphic only in the African sample and exhibiting an F_{ST} value within the particular range. The open portion shows the number polymorphic only in the non-African sample, while the gray portion indicates the number of sites polymorphic in both samples.

lescent simulations of samples of size 50 with 15 segregating sites and $4Nc = 20$, where N is the diploid population size and c is the recombination per generation (Hudson 1990). As in our data analysis from the *su(s)* and *su(w^a)* loci, only sites where the rarer state occurred at least twice in the sample were considered. The value 0.337 was obtained by fitting a and λ in $r^2 = 0.02 + a/(1 + \lambda * \text{distance})$ to the 45,434 simulated r^2 values, where *distance* is in units of $4Nc$ (Hill and Weir 1994). The fitted λ for the simulated data was 0.354. The impact of recombination (proportional to distance) can be extended to include random gene conversion events with exponentially distributed tract lengths,

$$r^2 = 0.02 + \frac{0.337}{1 + c \cdot \text{distance} + 2gt(1 - e^{-\text{distance}/t})},$$

where g is the rate of gene conversion, t is the average gene conversion tract length, and c is the rate of crossing over (Andolfatto and Nordborg 1998). These parameters (all constrained ≥ 0.0) were fitted by least-squares to the *su(s)* and *su(w^a)* r^2 values separately and the predicted curves are also plotted in Figure 5. Remarkably, the fitted curves for the linkage disequilibria at the two loci are quite similar. The fitted estimates of rate of crossing over, c, are both 0.0. The estimated rates of gene conversion, g, are similar, 0.010 [*su(s)*] and 0.006 [*su(w^a)*]; these would be scaled in $4N$ under equilibrium Wright-Fisher sampling in the absence of selection. And the estimated tract lengths, 302 and 538 bp, respectively, are also not significantly different.

Two aspects of these distributions of linkage disequilibria at *su(s)* and *su(w^a)* in the African sample are unexpected. First the scale of linkage disequilibrium is not different between these two loci, despite the fact that the density of crossing over per physical length is estimated to be as much as 10-fold less at *su(s)*. Equally surprising is the fact that this pattern and scale of linkage disequilibrium is the same as that observed at the *white* locus (Miyashita and Langley 1988; Miyashita *et al.* 1993). The *white* locus is in a genomic region of normal crossing over and exhibits much higher levels of polymorphism. It is clear that the rate of crossing over is not the determining parameter of linkage disequilibrium within these loci. Gene conversion is the obvious alternative recombination mechanism. The scale of strong linkage disequilibrium ($r^2 > 0.3$; <400 bp) is approximately that reported for both meiotic and mitotic gene conversion in Drosophila. While quantification of gene conversion rates in Drosophila is limited, the evidence indicates that half or more of all meiotic recombination events among intragenic sites are gene conversions (Finnerty 1976). Thus gene conversion should have at least the same impact on intragenic linkage disequilibrium as crossing over. Most meiotic and mitotic conversion tracts are small (<500), comparable to the estimates above. The reported mean *P*-element-associated conversion tract is ≈1400 bp (Preston and Engels 1996), while that for meiotic gene conversion tracts is estimated to be ≈400 bp (Hilliker *et al.* 1994).

From both the distributions of r^2 in Figure 5 and

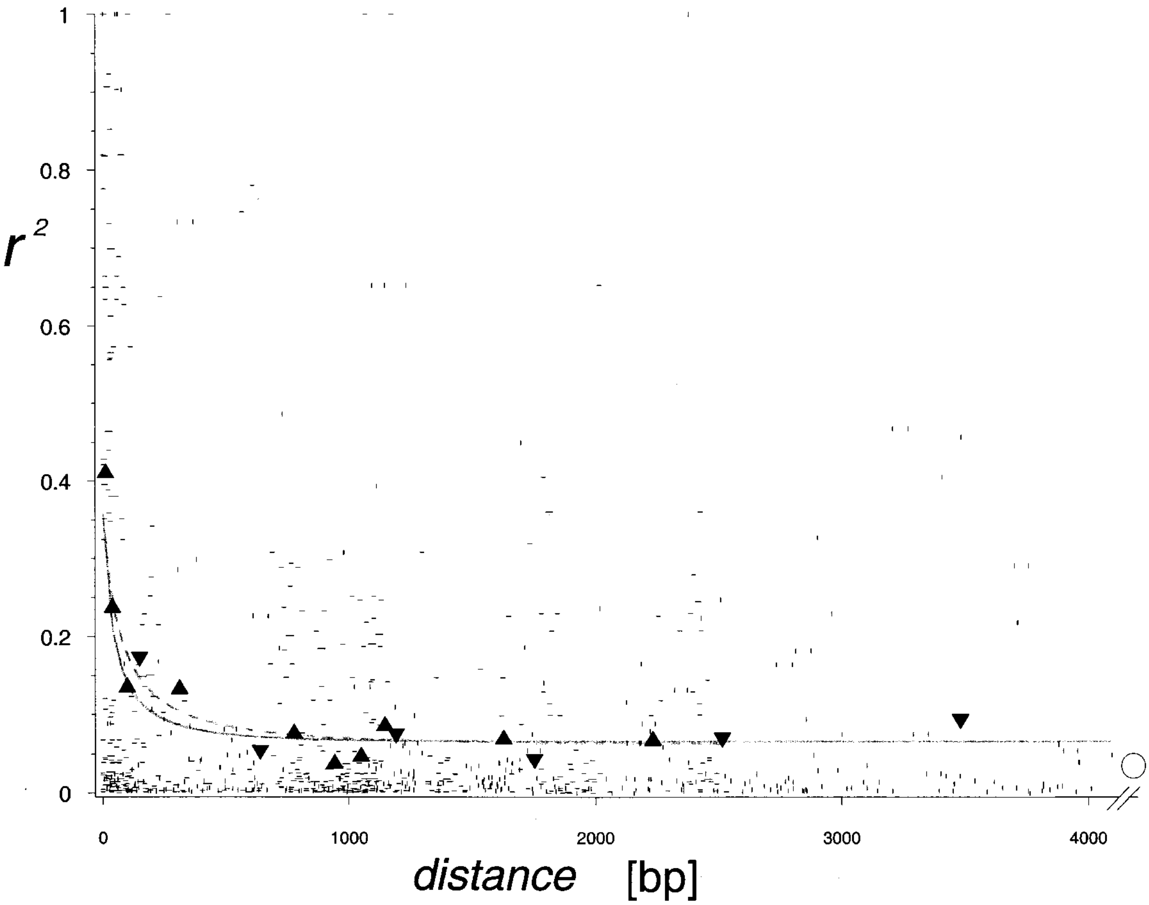

Figure 5.—Linkage disequilibria *vs.* genomic distance at *su(s)* and *su(w^a)* in the African sample. The squared correlation coefficient, r^2, for pairs of sites in *su(s)* (horizontal ticks) and in *su(w^a)* (vertical ticks) is plotted against the number of base pairs separating the sites. The mean r^2 over six contiguous intervals (*k* values) are also plotted: *su(s)*, ▲, $k = 63$; and *su(w^a)*, ▼, $k = 105$. Also plotted is a fit for each locus of the equation

$$r^2 = 0.02 + \frac{0.337}{1 + c \cdot \text{distance} + 2gt(1 - e^{-\text{distance}/t})};$$

dashed line, *su(s)*; and solid lines, *su(w^a)* (see text). The parameters, the rate of gene conversion, *g* (events per base pair per generation in the population), the mean length (base pairs) of a gene conversion tract, *t*, and the rate of crossing over per base pair, *c* (events per base pair per generation in the population), were estimated by nonlinear regression to the formula (see text): *su(s)*, $\hat{g} = 0.010$, $\hat{t} = 302$, and $\hat{c} = 0.0$; *su(w^a)*, $\hat{g} = 0.006$, $\hat{t} = 538$, and $\hat{c} = 0.0$. The mean r^2 for all the between-locus [*su(s)* − *su(w^a)*] pairs of polymorphic sites is plotted as a large circle at the far right.

the inferred *minimum* recombination event depicted in Figures 1 and 2, it is evident that these loci experienced a great deal of recombination in the history of these sampled alleles, despite their location in a region of greatly reduced crossing over and despite the drastic reduction in polymorphism (and thus time since the last common ancestor of the sampled alleles). No theoretical analysis of the impact of intragenic recombination (*e.g.*, gene conversion) on the background selection effect or on the directional hitchhiking effect has been reported. If the impact of background selection on linkage disequilibrium is similar to its effect on polymorphism, it can be approximated by $r^2 = 1/(1 + 4Nf_0c)$, where f_0 is the fraction of the deleterious-mutation-free chromosomes in the population, which can be estimated as the relative reduction in expected heterozygosity per nucleotide site (Charlesworth *et al.* 1993; Charlesworth 1994, 1996). The scale of linkage disequilibrium (in base pairs) should increase by $1/f_0$, even if gene conversion is the dominant source of intragenic recombination and if the rates of gene conversion are similar in regions of low and normal crossing over per physical length. Similar arguments and preliminary computer simulations extending our previous approach (Braverman *et al.* 1995) suggest that directional hitchhiking will also increase the scale of linkage disequilibrium, if gene conversion rates are uniform.

The observation of extensive recombination (presumably gene conversion) in the obviously short histories of these alleles at loci in regions of low crossing over per physical length (and low polymorphism) demands a modification or extension of the proposed linked-selected-perturbation models if they are to survive as general explanations of the correlation between standing polymorphism and crossing over per physical length. The rate of gene conversion could go up as the crossing over rate goes down. While there is no evidence to support this idea, it may be that the suppression of crossing over (near centromeres and telomeres) is simply the shunting of incipient crossovers toward gene conversions. Another interesting potential explanation for the high level of recombination in the histories of these alleles is a heterozygosity-dependent rate of gene conversion (see Stephan and Langley 1992). If gene conversion occurs at a higher rate between alleles that differ by fewer sites, the rate of recombination would change dynamically with increasing heterozygosity. Thus as directional hitchhiking or background selection reduced standing polymorphism, the recombination rate would increase. At equilibrium, regions of low crossing over would contain less polymorphism and thus undergo more gene conversion. This hypothesis of a heterozygosity-dependent rate of gene conversion is supported by observations in many systems including Drosophila. Nassif and Engels (1993) concluded that 0.5% sequence heterozygosity reduces the rate of *P*-element-induced mitotic gene conversion to 25% of that without heterozygosity. Variations in levels of heterozygosity have little or no effect on the rates of crossing over (Rutherford and Carpenter 1988). This is attributed to the strong regulatory processes underlying interference and the interchromosomal effect in crossing over (Lucchesi and Suzuki 1968). It is not known if heterozygosity-dependent inhibitions occur for gene conversion arising during meiosis or by other mechanisms in mitotic cells in Drosophila, but it is well known in yeast (see Kirkpatrick *et al.* 1998; Chen and Jinks-Robertson 1999). If this indication of population genetic consequences of the interaction between DNA sequence heterozygosity and rates of genetic exchange can be corroborated, our theories of the forces shaping genomic polymorphism and divergence will require interesting improvements.

We thank the many colleagues whose comments, questions, and criticisms have helped us improve this report. Special thanks to R. Hudson for the suggested extension to random conversion tract length in the fitted equation in the Figure 5 legend and to Peter Andolfatto for queries leading to better data quality. This work was supported by National Science Foundation (NSF) grant DEB 95-09548. E.H. was supported by a postdoctoral fellowship from the Finnish Research Council. J.M.B. was supported by a NSF/Sloan Foundation Postdoctoral Research Fellowship in Molecular Evolution and a postdoctoral research fellowship from the Ministerio de Educación y Ciencia, Spain.

LITERATURE CITED

Aguadé, M., and C. H. Langley, 1994 Polymorphism and divergence in regions of low recombination in *Drosophila*, pp. 67–76 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding. Chapman & Hall, New York.

Aguadé, M., N. Miyashita and C. H. Langley, 1989 Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122:** 607–615.

Aguadé, M., W. Meyers, A. D. Long and C. H. Langley, 1994 Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and su(*w^a*) regions of the *Drosophila melanogaster X* chromosome. Proc. Natl. Acad. Sci. USA **91:** 4658–4662.

Andolfatto, P., and M. Nordborg, 1998 The effect of gene conversion on intralocus associations (letter). Genetics **148:** 1397–1399.

Aquadro, C. F., D. J. Begun and E. C. Kindahl, 1994 Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46–56 in *Non-neutral Evolution: Theories and Molecular Data*, edited by B. Golding. Chapman & Hall, New York.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356:** 519–520.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Berry, A. J., J. W. Ajioka and M. Kreitman, 1991 Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129:** 1111–1117.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

Cavalli-Sforza, L. L., P. Menozzi and A. Piazza, 1994 *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.

Charlesworth, B., 1994 The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63:** 213–227.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68:** 131–149.

Charlesworth, B., 1998 Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. **15:** 538–543.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Chen, W., and S. Jinks-Robertson, 1999 The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast. Genetics **151:** 1299–1313.

David, J. R., and P. Capy, 1988 Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4:** 106–111.

Dvorák, J., M. C. Luo and Z. L. Yang, 1998 Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics **148:** 423–434.

Finnerty, V., 1976 Gene conversion in *Drosophila*, pp. 331–349 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner and E. Novitski. Academic Press, London/New York.

Gillespie, J. H., 1997 Junk ain't what junk does: neutral alleles in a selected context. Gene **205:** 291–299.

Hill, W. G., and B. S. Weir, 1994 Maximum-likelihood estimation of gene location by linkage disequilibrium. Am. J. Hum. Genet. **54:** 705–714 (Erratum, Am. J. Hum. Genet. **55**(1): 217).

Hilliker, A. J., G. Harauz, A. G. Reaume, M. Gray, S. H. Clark *et al.*, 1994 Meiotic gene conversion tract length distribution within the *rosy* locus of *Drosophila melanogaster*. Genetics **137:** 1019–1026.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–44 in *Oxford Series in Evolutionary Biology*, edited by D. Futuyma and J. Antonovics. Oxford University Press, New York.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Hudson, R. R., M. Slatkin and W. P. Maddison, 1992 Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583–589.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

Kirkpatrick, D. T., M. Dominska and T. D. Petes, 1998 Conversion-type and restoration-type repair of DNA mismatches formed during meiotic recombination in *Saccharomyces cerevisiae*. Genetics **149:** 1693–1705.

Langley, C. H., J. MacDonald, N. Miyashita and M. Aguadé, 1993 Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region in *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90:** 1800–1803.

Lucchesi, J. C., and D. T. Suzuki, 1968 The interchromosomal control of recombination. Annu. Rev. Genet. **2:** 53–86.

Martín-Campos, J. M., J. M. Comerón, N. Miyashita and M. Aguadé, 1992 Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130:** 805–816.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of favorable genes. Genet. Res. **23:** 23–35.

Miyashita, N., and C. H. Langley, 1988 Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics **120:** 199–212.

Miyashita, N. T., M. Aguadé and C. H. Langley, 1993 Linkage disequilibrium in the *white* locus region of *Drosophila melanogaster*. Genet. Res. **62:** 101–109.

Nachman, M. W., 1997 Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147:** 1303–1316.

Nassif, N., and W. Engels, 1993 DNA homology requirements for mitotic gap repair in Drosophila. Proc. Natl. Acad. Sci. USA **90:** 1262–1266.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. **67:** 159–174.

Preston, C. R., and W. R. Engels, 1996 P-element-induced male recombination and gene conversion in *Drosophila*. Genetics **144:** 1611–1622.

Rutherford, S. L., and A. T. Carpenter, 1988 The effect of sequence homozygosity on the frequency of X-chromosomal exchange in *Drosophila melanogaster* females. Genetics **120:** 725–732.

Stephan, W., 1994 Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae*, pp. 57–66 in *Non-Neutral Evolution*, edited by B. Golding. Chapman & Hall, New York.

Stephan, W., and C. H. Langley, 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

Stephan, W., and C. H. Langley, 1992 Evolutionary consequences of DNA mismatch inhibited repair opportunity. Genetics **132:** 567–574.

Stephan, W., and C. H. Langley, 1998 DNA polymorphism in *Lycopersicon* and crossing over per physical length. Genetics **150:** 1585–1593.

Stephan, W., L. Xing, D. A. Kirby and J. M. Braverman, 1998 A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. Proc. Natl. Acad. Sci. USA **95:** 5649–5654.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Communicating editor: R. R. Hudson